

Teacher's feedback vs. computer-generated feedback: A focus on articles

Tamara Hernández Puertas

tamaraeoi@gmail.com

Escuela Oficial de Idiomas (Castellón), Spain

ABSTRACT

As attested by a vast number of studies, in the process of second/foreign language acquisition feedback plays an important role as it may trigger learners' noticing of the mismatch between their interlanguage and the target language (Schmidt 1990). In foreign language classrooms, feedback on written production may not be properly provided due to a large number of students or time constraints (Chacón-Beltrán 2017). In this sense, the use of new technologies in the classroom may help both the teacher in the correction process and the student in his/her language development. In the present study we aim to compare feedback provided by the teacher and feedback provided by the software Grammar Checker (Lawley 2015). One group of English-as-a-foreign language (EFL) students received teacher's feedback on their mistakes on articles in their written production whereas a second group obtained feedback on the same grammar aspect by means of the above-mentioned software. The control group did not obtain feedback on their errors. Results show statistically significant differences in the last composition for the group who received teacher's feedback, although this feedback did not have a lasting effect in the tailor-made delayed test. In light of these findings, we may claim that the use of Grammar Checker as a potential tool for self-correction and feedback may facilitate students' language development, at least on the grammar aspect under analysis.

Keywords: *corrective feedback, teacher's feedback, computer-generated feedback, writing, articles, errors*

I. INTRODUCTION

Second language acquisition (SLA) is a complex process involving multiple variables along with natural elements such as errors, which should be regarded as part of the language learning process and not as something negative that has to be avoided. By means of errors, learners may test their hypotheses about how the target language works and teachers obtain information about learners' progress and difficulties in their development. Traditionally, teachers (and sometimes, peers) have provided correction in the formal context in various ways to help learners overcome their errors (both oral and written) and further their learning. The issue of whether mistakes should be corrected, when and how, among other questions, has fuelled much research, together with the elaboration of different typologies accounting for corrective feedback (CF) types,

ranging from most indirect to most direct. However, there seems to be some agreement on the fact that, although demanded by the learners, providing CF is a complex task to do. Corrective feedback for oral mistakes may be obtrusive and thus interrupt the flow of conversation. In turn, CF for written errors may take much of the teacher's time and sometimes it is only provided superficially.

Over the past two decades, there have been efforts to develop software which aids in the process of student writing along with some other software which provides a score on students' written production. The focus of this study is on the former, that is, we aim at contributing to the expanding body of research on computer-generated feedback in an attempt to examine whether this type of feedback has an impact on students' linguistic accuracy when compared to teacher's feedback. With this aim in mind, the software Grammar Checker was employed by one group of students as source of feedback on errors, whereas another group obtained teacher's feedback.

II. CORRECTIVE FEEDBACK AND SLA

Making mistakes is part of the natural process of learning a language. However, when producing output, students may not be aware of how successful they have been at conveying their messages if some kind of feedback is not offered. Corrective feedback becomes, then, a key factor in the SLA process since mere language exposure does not seem to be enough and second language (L2) speakers need some kind of corrective feedback to notice the discrepancies between their output and the L2.

The term corrective feedback (Lyster 1998) has adopted different terminology depending on the author: for example, 'negative evidence' (Long 1991), 'interactional feedback' (Lyster and Mori 2006) or 'negative feedback' (Ortega 2009). For the purposes of the present study, we will adhere to the definition provided by Russell and Spada (2006: 134): 'Corrective feedback will refer to any feedback, provided to a learner, from any source, that contains evidence of learner error of language form'. In this sense, corrective feedback refers to the teacher's reaction to a mistake, when this reaction causes attention to language forms and has a corrective aim. Much research has been carried out on CF, and most has employed different types of CF based on the learner's reaction (i.e., uptake). For instance, Ellis (2009) classified CF types along the

implicit-explicit dichotomy. Implicit feedback referred to recasts (i.e., reformulation of the learner's incorrect utterance minus the error), repetition and clarification requests, in which the learner has to work harder in order to spot the mistake and self-repair. In turn, explicit feedback included explicit corrections, metalinguistic explanations, elicitations and paralinguistic signals which showed in a more direct way that the learner's production was wrong.

Although the effectiveness of CF on acquisition is a debatable issue, it is regarded as an intervening element in the process of SLA. In fact, since the early 90s, a vast number of studies have demonstrated the beneficial role of CF on acquisition. Moreover, some meta-analyses and reviews of the literature (for example, Russell and Spada 2006, Spada 2011), point to the positive effects of CF for L2 grammar learning and its durability over time as long as it is noticeable, comprehensible and as individualized as possible.

II.1. The effect of corrective feedback on written production

In the current multimedia age, different modes of writing and image combine to make multimodal texts which communicate meanings and may be used for language learning. Images (including the use of colors) play an essential role in multimodal communication as attention-getters (Kress 2010), therefore maximizing the potential for learning. In this sense, a crucial condition for the effectiveness of CF is that the student notices the input features and the differences between his/her interlanguage and the target language forms. The notion of noticing was coined by Schmidt (1990) and supported by other researchers (e.g., Mackey et al. 2000, Philp 2003) as one of the crucial elements necessary for acquisition to take place, in the sense that noticing is essential for input to become intake. Intake has been defined by Ellis (1994: 708) as 'that portion of input that learners notice and therefore take into temporary memory'. Learners may notice input thanks to the CF provided to them in the language classroom. Indeed, research has shown that CF does occur in the classroom in a high proportion (e.g., Panova and Lyster 2002) as an intervening variable in the process of language learning. The benefits of CF in oral interaction point to learners' noticing of problematic forms, opportunities to modify output and test hypotheses, and an increase in linguistic

accuracy. Yet, the debate about the value of written corrective feedback (WCF, henceforth) has yielded conflicting results (Evans et al. 2010). For instance, in a much-cited study by Truscott (1996), he argued that ‘correction is not only unhelpful but even counterproductive’ (1996: 354). In the same vein, Polio et al. (1998) and Fazio (2001) stated that CF can be discouraging and ineffective to improve subsequent writings due to the pressure it may create on learners. However, broadly speaking, research has found a beneficial effect of CF on writing accuracy (e.g., Bitchener 2008, Lee 2013). More specifically, Bitchener and Ferris (2012) claim that students' accuracy improves when they attend to feedback as they draw their attention to linguistic inconsistencies or mistakes. Moreover, for ethical reasons, learners need to be provided with CF in their written production, even more when it has been shown that students want to improve their linguistic accuracy (Ferris and Hedgcock 2005) and that they expect to have their writing mistakes marked (Guénette 2007).

Feedback may be delivered in a more direct (explicit) or indirect (implicit) way. *Direct feedback* is offered when the teacher provides the correct form straight away and the student is supposed to incorporate that correction in the final version. Contrarily, in *indirect feedback* the teacher merely indicates in some way (underlining or highlighting the error, or marking in the margin of the text) that there is an error, without providing the correction. Thus, the student knows there is a mistake and he/she has to solve it. In this sense, some voices have claimed that indirect feedback is more desirable because it may engage students in problem solving and, eventually, in more progress in accuracy over time than direct feedback (Ferris et al. 2000). Different degrees of explicitness in feedback provision were examined in Ferris and Roberts' (2001) study: Group A had their errors underlined and coded, Group B had their errors underlined but not coded and Group C (control group) had no error markings. No statistically significant differences were reported between Group A and B, suggesting that more explicit feedback (underlining and coding of errors) was not more advantageous than simple underlining.

Some research has addressed the impact of different types of feedback on accuracy in student writing. Chandler (2003) had four treatments including (i) *Correction*, (ii) *Underlining with Description*, (iii) *Description of error only*, and (iv) *Underline*. Findings show that conditions (i) and (iv) resulted in more accurate pieces of writing in

the next assignment, whereas treatments (ii) and (iii), which involved a description of the error type, had the opposite effect. Overall, the number of studies which have addressed the effectiveness of direct and indirect WCF show inconclusive findings. However, there seems to be a wider consensus on the fact that if feedback is provided, learners' accuracy tends to improve when compared to control groups receiving no feedback, as reported by Ene and Upton's (2014) study.

II.2. Computer-generated feedback in writing

When to provide feedback has been one of the main concerns in the field of language correction and feedback. Warschauer (2010) claimed that autonomous learning and revision could be enhanced by promptly delivered feedback. Indeed, when little time lapses between the student's writing and the teacher's CF, learning opportunities may be maximized. In the same line, Guichon et al. (2012) argued that if learners can get 'just in time' feedback, they may self-correct almost immediately after their mistakes and possibly incorporate this feedback in subsequent writings. In this way, written CF may be more effective as in traditional classrooms feedback is only provided by the teacher several days after the written production.

As stated by Spada (2011), corrective feedback occurs both in natural learning contexts as well as in formal environments, although it is more frequent and presumably more beneficial and necessary in the latter. Yet, in large classes in which the students are required to perform written tasks, teachers need to lessen their workload by delegating work to their students, who may use electronic feedback to self-correct their written productions (Lee 2013). Therefore, more time could be devoted to other areas which need more attention in writing, such as content and organization (Chen and Cheng 2008). In this sense, and especially in the education domain, the importance of technology and the benefits it may provide to the learning process shows how it is taking over classrooms at all levels. The use of computer tools, what is called 'computer-assisted language learning' (CALL), applied to the classroom and the students' way of working represents an extra value and motivation. In fact, as Becker stated (1991: 385), 'in the 1980s, no single medium of instruction or object of instructional attention produced as much excitement in the conduct of elementary and

secondary education as did the computer.’ CALL is an approach that has many advantages: first, it adapts to the learning of the students letting them control their own pace, second, it allows them to be more autonomous since they are the ones who make their own choices, third, it offers them freedom and authenticity and finally, it develops their critical thinking. In this vein, computer-mediated feedback may contribute to help students write more independently both inside and outside the classroom. Moreover, research from Tiene and Luft (2001) suggests that the use of technology fosters individualized communication between teacher and students more often and allows teachers to focus on higher-order aspects of writing, leaving common grammar or spelling mistakes to the program.

As just stated, new technological implementations in the language classroom have influenced the skill of writing, especially the revision and editing processes by means of online tools. The interplay of range of modes on screen (for example, image and writing) has resulted in a redesign of how students can receive feedback. As Jewitt put it (2002: 172), ‘communication and learning are multimodal’. This multimodality may be significant for writing improvement. In this sense, in the past twenty years, software aiming at scoring and/or providing feedback on students’ writings has been devised (e.g., *Criterion*, *MyAccess*, *Grammarly*, *Summary Street*, to mention but a few), with diverse degree of effectiveness on students’ satisfaction (Chen and Cheng 2006). Still, some voices (e.g., Ware and Warschauer 2005) claim that the amount of time a teacher may spend correcting students’ compositions may be dramatically reduced if teachers can rely on computer-generated feedback. Moreover, software which generates feedback on writing has been created providing either reports on grammatical errors or more holistic assessment on aspects such as content or organization of the piece of writing. In the case of grammar checkers, Potter and Fuller (2008) reported an increase in students’ motivation, proficiency and confidence in grammar rules in the use of English grammar checkers. In turn, Nadasdi and Sinclair (2007) argued that the French online grammar checking program *BonPatron* was as effective as teacher correction. Also, Burston (2008) investigated the accuracy of this grammar checker showing that 88% of errors were spotted by the software. Mistakes were highlighted by means of color-coding: red indicated those grammatical aspects the student had to modify and orange was used to signal segments or words which needed to be verified.

Despite the a priori benefits of grammar checkers, they are not without limitations. As argued by Davis (1989), any user of grammar checkers has to set their perceived usefulness and ease of use, two key factors in Davis' Technology Acceptance Model (TAM) determining the likelihood of acceptance of new technology. A second drawback refers to the fact that sometimes computer-generated feedback may not be specific or informative enough to guide learners in their revision process, eventually causing frustration or dissatisfaction (Chen and Cheng 2006).

III. GRAMMAR CHECKER

In 2001 the Universidad Nacional de Educación a Distancia (UNED) in Madrid started to work on the software Grammar Checker (GC, henceforth) in an attempt to detect errors made by English-as-a-foreign-language students. It provides written feedback on grammar, spelling, and words used incorrectly based on a corpus of eighty million words 'taken from the written component of the British National Corpus' (Lawley 2015: 26). As explained by this author, the program divides the text into segments that are compared to that corpus and highlighted in red if they do not appear in it or have a threshold number lower than 0 and 0.1, in orange if they occur in the corpus fewer than 500 times and their threshold numbers range between 0.1 and 0.5, or yellow if they occur fewer than 75 times and their threshold numbers lie between 0.5 and 0.9. Therefore, this program requires cognitive process from students as it only uses certain colors to show frequency but does not offer the possibility to receive corrections at the click of a mouse. Students are responsible for changing the segment or not upon reflection. In this way, it offers the opportunity to learn from mistake. GC does not provide a score for the text, it merely alerts users to those combinations that are rare or do not occur.

GC works as follows: after creating an account, the student has to write the text and press "Enter your text" and then "Start" to check if there are any mistakes. First, spelling mistakes are highlighted in yellow (also purple if it is a very rare word but not necessarily a mistake, e.g., proper names) and by clicking on the words highlighted useful feedback is provided. By clicking on "Modify", the previous spelling mistakes can be corrected and checked again by pressing "Check again". Then the same

procedure is followed for the “Incorrect sequences filter” that highlights grammar mistakes such as ‘These table’, and for the “Problem words filter” which refers to correct English and does not highlight any word but suggests words that are usually misused by students, e.g., ‘insano’ (unhealthy). Therefore, if after reading the suggestions the student thinks he/she has made a mistake, he/she can modify it.

The most important step for the aims of the present study is the button “Pairs filters” which highlights phrases that do not usually occur, e.g., ‘had do’. In order to know the frequency with which those phrases occur and decide whether it is a mistake or not, the student can use the search engine at the top of the screen. Figure 1 below illustrates a screenshot of GC:

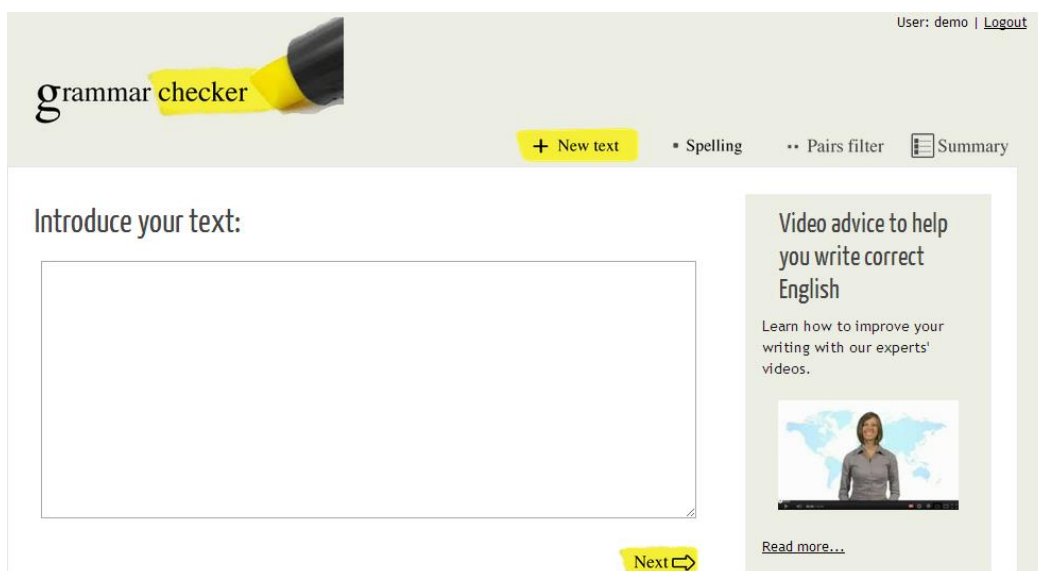


Figure 1. Screenshot of Grammar Checker.

GC was selected for the purposes of the present study for several reasons: firstly, it offers a cue (highlighting in colors) so that students can locate, reflect and self-correct, which, according to the literature, may be conducive to learning. Secondly, GC does not overwhelm language learners with metalinguistic terminology which may be at odds with some learners’ literacy (Dikli 2006). Thirdly, this software does not score learner’s written production, but provides them with feedback and possible suggestions for improvement. Finally, it is an affordable program for only €14 a year for students

aiming, in the present study, for level B2 of the Common European Framework of Reference of Languages (CEFR).

IV. THE STUDY

Prior to this research, a pilot study to test the use of GC was conducted with a group of students with a similar level of proficiency as the participants in the present study and also enrolled in an Official School of Languages. The purpose of that pilot study was, on the one hand, to test the computer program, and on the other hand, to decide important aspects such as the level and the number of students participating and the targeted grammatical aspects (articles, verb tenses and prepositions in this case). One group of students received teacher's feedback and another obtained feedback by means of GC. Analysis of the data collected in the pilot study revealed a higher number of corrections after computer feedback. Therefore, this program proved helpful in highlighting and correcting students' mistakes.

Taking into account previous research pointing at overall benefits of WCF in the development of students' writing accuracy on the one hand (Bitchener 2008, Russell and Spada 2006), and the rapid growth of computerized feedback in educational contexts on the other (Ene and Upton 2014), in this study we entertain two research questions. The first research question aims at revealing what type of feedback (teacher vs. computer) will have a better effect on accuracy in the targeted grammar aspect (articles). On the other hand, the second research question aims at showing what type of feedback (teacher vs. computer) will have a lasting effect in the delayed tailor-made test.

IV.1. Participants

Three groups of Spanish students (n=27) participated in the present study. They were divided into two treatment groups and the control group. All participants were studying at an Official School of Languages in order to pass the B2 level for professional reasons and reported having studied English for over 6 years. Their mother tongue was Spanish and/or Catalan and their ages ranged from 20 to 50 years old (mean=39.3).

The study was carried out as part of their formal EFL instruction and the compositions were regular assignments the students had to elaborate as part of their written homework.

IV.2. Targeted grammar aspect: articles

Errors on rule-governed forms allow for more focused correction than errors which are not rule-based (Lee 2013). Ferris (1999) termed the first type of errors ‘treatable errors’, as some grammar errors may be treatable through feedback. In this vein, articles fall under this ‘treatable’ category and for Spanish EFL students they may be a recurrent source of errors, especially the zero article. In fact, the English article system has been shown to be used inaccurately by foreign language learners, even with high proficiency. Despite the fact that article errors seldom cause misunderstanding, since they possess low communicative value, it is still necessary for learners to overcome their problems with this specific grammar form. On this account, Master (1995) pointed out that attention to the article system was important because this type of errors may leave the impression that the learners have incomplete control of the target language. Some years later, Bitchener (2008) also argued that EFL learners across different language proficiency levels experience difficulties in their mastery of the English article system. These perceived difficulties, along with the fact that articles are potentially ‘treatable’, were the reasons to have articles as targeted grammatical forms for examination.

IV.3. Types of feedback

Group 1 (n=11) received teacher’s feedback, Group 2 (n=8) computer feedback and the Control Group (n= 8) obtained no feedback on the targeted grammatical aspect.

Computer feedback was provided by Grammar Checker by means of a color code (red, orange and yellow) as explained in Section III. It was an indirect type of feedback which only signaled potentially problematic bits in the compositions. For comparability issues, teacher’s feedback had to be indirect as well, so she also used colors similar to the ones in the computer software to highlight the mistakes on articles.

IV.4. Data collection procedure

In a session prior to the data collection, participants belonging to Group 2 were trained in the use of *Grammar Checker* and they were explained what the color code meant and how they had to correct their mistakes. Afterwards, all participants were asked to write a 180/200-word composition based on a comic strip (*Abbey Time 1*). In strip 1 someone is writing a letter to an old woman, in strip 2 Abbey appears next to an Elvis-looking man, in strip 3 the man is holding some flowers and a teddy bear, in strip 4 a woman different from the old woman and physically similar to Abbey is looking at the man with a menacing gaze, in strip 5 Abbey looks sad and in strip 6 someone who seems to be Abbey is writing a letter. As mentioned above, Group 1 received teacher's feedback and Group 2 obtained computer feedback. The control group did not get any feedback on articles but on other non-targeted grammatical aspects. After this feedback, they rewrote a second version of the same comic strip (*Abbey Time 2*) to check whether correction had been effective. The time elapsed between Abbey 1 and teacher's feedback was one week, and between teacher's feedback and Abbey 2 also one week.

Two weeks after Abbey 2, participants composed a second text based again on a similar graphic prompt but with different strips (*Pam Time 1*). In strip 1 someone is writing a letter while the image of Pam appears in the background. In strip 2 an old woman is holding a sheet of paper, and in strip 3 the woman who looks like Pam is looking at the Elvis-looking man with a menacing gaze. In strip 4 the man is showing the woman a cake he has just made, in strip 5 the old woman looks happy and in strip 6 the old woman is writing a letter.

The same process as the one depicted above applied: after the first composition (*Pam Time 1*), feedback (either by the computer software or the teacher) was provided and students wrote a second version (*Pam Time 2*) after 2 weeks from the first version. Therefore, 4 compositions (Abbey Time 1 and 2 and Pam Time 1 and 2) are the data for analysis.

Six weeks after having written the last of the four compositions, the participants were asked to complete an individual tailor-made test (see a sample in Appendix 1) to check any long-term impact of the two types of feedback. The tailor-made tests included all the errors each student had made in Abbey Time 2 and Pam Time 2, that is, after having

obtained feedback three times (either from the teacher or the computer). Table 1 illustrates the timeline for data collection.

Table 1. Timeline for the data collection procedure.

Week 1	Abbey T1
Week 2	Teacher's or computer feedback
Week 3	Abbey T2
Week 4	Teacher's or computer feedback
Week 5	Pam T1
Week 6	Teacher's or computer feedback
Week 7	Pam T2
Week 8	Teacher's or computer feedback
Week 14	Tailor-made test

All four compositions belonged to the same genre, that of narrative story, in which a short story is described. The learners had to describe what was happening in the story according to the given pictures. Therefore, as stated by Bitchener (2008), valid text comparisons can be made because both storylines were related and even seemed a continuation and had similar characters. For this reason, similar tenses, structures and vocabulary for both comic strips were expected.

IV.5. Results and discussion

A Kruskal-Wallisⁱ test was run to determine whether there existed significant differences in the two experimental groups and the control group taking into account errors on articles in Abbey Time 1, that is, in the first composition the learners had to write. As can be seen in Table 2, results show no significant differences, a fact that, from a methodological point of view, is desirable as it indicates that all groups made an equivalent number of errors ($p > 0.05$ in all three groups).

Table 2. Means and standard deviations for Abbey Time 1.

Group	Mean and standard deviation
Group 1: computer's feedback	.91 (2.07)

Group 2: teacher's feedback	1.13 (1.35)
Control group	.50 (.75)

As to the first research question, a first analysis was carried out to determine whether feedback had been useful when students had to write Abbey Time 2 and Pam Time 2 (i.e., when they had obtained feedback after Abbey Time 1 and Pam Time 1). With that aim in mind, a Wilcoxon signed-rank testⁱⁱ taking into account the number of errors on articles between Abbey Time 1 and 2, and between Pam Time 1 and 2 revealed only statistically significant differences between Pam 1 and 2 for the group who had been offered teacher's feedback (Group 1; $p=.026$). For Group 2 (computer group) and the Control Group, no significant differences were observed, as Table 3 depicts:

Table 3. Comparison between Time 1 and Time 2 in both compositions.

	Group 1 (teacher)	Group 2 (computer)	Control Group
	$Z_{(w)}$	$Z_{(w)}$	$Z_{(w)}$
Abbey Time 1 and 2	1.00	.00	.81
Pam Time 1 and 2	2.23	.68	.33

As stated above, only a significant decrease in the number of errors in the use of articles occurs between Pam 1 and 2 for Group 1. Although both treatment groups at the time of writing Pam 2 had received feedback three times, in light of our results teacher's feedback appears to be more effective as far as linguistic accuracy is concerned, despite the fact that this feedback was as indirect as the one provided by the computer. In view of the above results, the effect size was calculated (Cohen's d^{ii}). For Pam Time 1 and 2, the effect size was large ($d=1.024$), but the rest of effect sizes ranged from medium to small.

A second test was used (Wilcoxon signed-rank test) to examine the effect of feedback in Abbey and Pam at Time 2. Again, as shown in Table 4 below, the analysis reveals only statistically significant differences for Group 1, that is, it seems that teacher's feedback had a positive effect on reducing learners' errors on articles. One possible explanation for this finding is that learners tried harder to self-repair before giving their revised compositions back to their teacher. Maybe they were not so confident about computer's

feedback and might have felt skeptical about this source of feedback. Still, that is the only significant difference, since Group 2 and the CG did not show any significant difference in reducing the number of errors. Our results seem to align with Sauro's (2009) research on zero articles. Her two treatment groups received two types of computer feedback (recast and metalinguistic information). The indirect type of feedback (recast) in Sauro's study and highlighting in the present investigation do not seem to have an impact on learners' correction of their errors.

Table 4. Means and standard deviations for Abbey Time 1.

Abbey & Pam Time 2	$Z_{(w)}$	p
Group 1 (computer)	2.11	.035
Group 2 (teacher)	.37	.70
Control Group	.81	.41

In an overview of the grammar checker *Grammarly*, Cavaleri and Dianati (2016) report that 22% of their students agreed that the feedback provided on their writing was not always helpful, as some of the feedback made no sense for learners. Our participants may presumably have been in the same situation, finding the feedback too indirect.

As for the second research question, a Wilcoxon test was run. In Group 1, there were no statistically significant differences ($Z_{(w)} = 1.63$; $p = .10$; $d = .25$) between the errors students had made in Abbey Time 2 and Pam Time 2 and the tailor-made tests, showing a small effect (calculated with Cohen's d). The same pattern applies to the results for the computer group and the control group, as there were no significant differences between the mistakes in Time 2 in both compositions and the tailor-made tests ($Z_{(w)} = 1.63$; $p = .10$; $d = .14$) for Group 2 and ($Z_{(w)} = 1.89$; $p = .059$; $d = .40$) for the CG, again with a small to medium effect size.

Despite the fact that, as shown by the results of the first research question, there were significant differences in the number of errors after teacher's feedback, this applied only to immediate gains which were not maintained in the long term, as attested by the results for the second research question. Neither of the treatment groups showed gains in accuracy in the tailor-made post-tests. Again, one likely explanation for this result be

the fact that feedback was too indirect and the color codes were too vague and not showing the learners what to focus on in a more specific way. In this sense, multimodal combination of text and image (colors, in this case) did not seem to benefit the students' self-correction process. Although it has been claimed that learners may benefit more from indirect CF because they need to engage in deeper language processing (van Beuningen et al. 2008), CF which is too indirect may not reach the desired goals in the long run. Indeed, Chandler (2003) found that direct feedback resulted in largest accuracy gains, both in revisions of previous writings and in subsequent writing, whereas students who revised their compositions after indirect CF were unable to do so.

A second explanation points to the fact that the compositions learners had to write were not graded. As a result, their motivation could have been rather low along with the possibility that they might have got bored of writing four compositions which were very similar and demanded little creativity.

V. CONCLUSION

Many adult students may have to work autonomously on their language acquisition process. As shown by the findings of the present study, computer-assisted learning tools such as *Grammar Checker* may prove useful in that process, as 'everything that can be done to facilitate accurate self-correction is positive' (Lawley 2016: 879). Still, GC merely suggests potential problems by highlighting some written bits, thus leaving it up to students to solve the error. In this vein, computer-generated feedback may have resulted to be a difficult task for the students who received this type of feedback 'due to their learned dependence on teacher-provided feedback' (Peterson 2017: 48). Moreover, the effectiveness of computer-generated feedback to highlight aspects such as content or organization of writings is questionable as humans can assess writings more accurately than computers (Reiners et al. 2011).

The present study aimed at comparing the impact of teacher's and computer feedback on students' errors, as most errors are repeated among students, which makes the teacher correct the same error numerous times. In this sense, and despite the above-mentioned drawbacks of using technology for grammar correction, software such as

Grammar Checker could improve this situation, encouraging students to be more independent of the teacher and more responsible for their own learning. Benefits may apply both for the learners and the teacher.

Yet, taking into account the results of this study, we concur with Ware's (2011) claims that computer-generated feedback should be seen as a supplement to writing instruction and not as a replacement, since teacher's CF, although as indirect as the one delivered by GC, seemed to work better in reducing the number of errors in the short run. We adhere to Heift and Hegelheimer' (2017) recent claims that there is still scant evidence with regard to whether computer-generated feedback results in accuracy development and learning over time, pointing to a need of long-term research to determine these issues.

This piece of research was conducted in authentic classrooms as part of students' ordinary classes. In this sense, it represents a realistic picture of EFL instruction, which impacts on its ecological validity, even though some factors, such as students' commitment during the process may be a handicap. Therefore, as limitations to the study we can mention the small sample size, which poses questions of generalizability, and the fact that the feedback provided addressed errors on articles, that is, rule-governed forms which are more amenable to correction (Lee 2013). The extent to which other non-rule-governed aspects may benefit from the two types of CF has not been examined in the present study. Also, the type of indirect feedback offered (highlighting errors) may prove more useful for students at higher levels of proficiency. Perhaps the small impact of this kind of feedback in the present study may be due to the proficiency level of the participants, who could have felt at a loss because of their limited linguistic competence. Finally, a further limitation refers to the effectiveness of *Grammar Checker*, since it depends highly on the teacher and students' attitudes toward computer-based feedback and their technology-use skills in working with computer-based programs, because not all teachers and students may be equally skilled.

Notes

ⁱ Non-parametric test that compares independent sample of equal of different sample sizes.

ⁱⁱ Non-parametric test used to compare two related samples in this case

REFERENCES

- Becker, H. J.** 1991. "How computers are used in United States schools: Basic data from the 1989 IEA Computers in Education survey". *Journal of Educational Computing Research*, 7, 385-406.
- Bitchener, J.** 2008. "Evidence in support of written corrective feedback". *Journal of Second Language Writing*, 17 (2), 102-118.
- Bitchener, J. and Ferris, D. R.** 2012. *Written corrective feedback in second language acquisition and writing*. New York: Routledge.
- Burston, J.** 2008. "BonPatron: An online spelling, grammar, and expression checker". *Calico Journal*, 25 (2), 337-347.
- Cavaleri, M. and Dianati, S.** 2016. "You want me to check your grammar again? The usefulness of an online grammar checker as perceived by students". *Journal of Academic Language & Learning*, 10 (1), 223-236.
- Chacón-Beltrán, R.** 2017. "Free-form writing: computerized feedback for self-correction". *ELT Journal*, 71 (2), 141-149.
- Chandler, J.** 2003. "The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing". *Journal of Second Language Writing*, 12, 267-296.
- Chen, C-F. and Cheng, W-Y.** 2006. *The use of computer-based writing program: facilitation of frustration?* Paper presented at the 23rd. International Conference on English Teaching and Learning in the Republic of China.
- Chen, C-F. and Cheng, W-Y.** 2008. "Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes". *Language Learning & Technology*, 12 (2), 94-112.
- Davis, F. D.** 1989. "Perceived usefulness, perceived ease of use, and user acceptance of information technology". *MIS Quarterly*, 13 (3), 319-339.
- Dikli, S.** 2006. "An overview of automated scoring of essays". *Journal of Technology, Learning, and Assessment*, 5 (1), 1-35.

- Ellis, R.** 1994. *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Ellis, R.** 2009. "A typology of written corrective feedback types". *ELT Journal*, 63 (2), 97-107.
- Ene, E. and Upton, T. A.** 2014. "Learner uptake of teacher electronic feedback in ESL composition". *System*, 46, 80-95.
- Evans, N. W., Hartshorn, K. J., McCollum, R. M. and Wolfersberger, M.** 2010. "Contextualizing corrective feedback in second language writing pedagogy". *Language Teaching Research*, 14 (4), 445-463.
- Fazio, L.** 2001. "The effect of corrections and commentaries on the journal writing accuracy of minority- and majority-language students". *Journal of Second Language Writing*, 10 (4), 235-249.
- Ferris, D. R.** 1999. "The case for grammar correction in L2 writing classes: A response to Truscott (1996)". *Journal of Second Language Writing*, 8 (1), 1-11.
- Ferris, D. R., Chaney, S. J., Komura, K., Roberts, B. J. and McKee, S.** 2000. *Perspectives, problems, and practices in treating written error*. Colloquium presented at International TESOL Convention. (March 14-18, 2000).
- Ferris, D. R. and Hedgcock, J.** 2005. *Teaching ESL composition: Purpose, process, and practice*. Mahwah, NJ: Erlbaum.
- Ferris, D. R. and Roberts, B.** 2001. "Error feedback in L2 writing classes. How explicit does it need to be?" *Journal of Second Language Writing*, 10, 161-184.
- Grammar Checker.** 2 October 2015. <http://www.e-uned.es/subscription/subscriptionsInfo.php?subID=CM>
- Guénette, D.** 2007. "Is feedback pedagogically correct? Research design issues in studies of feedback on writing". *Journal of Second Language Writing*, 16, 40-53.
- Guichon, N. Betrancourt, M. and Prié, Y.** 2012. "Managing written and oral negative feedback in a synchronous online teaching situation". *Computer Assisted Language Learning*, 25 (2), 181-197.

- Heift, T. and Hegelheimer, V.** 2017. "Computer-assisted corrective feedback and language learning". In Nassaji, H. and Kartchava, E. (Eds.) *Corrective feedback in second language teaching and learning*. New York: Routledge, 51-65.
- Jewitt, C.** 2002. "The move from page to screen: The multimodal reshaping of school English". *Journal of Visual Communication*, 1 (2), 171-196.
- Kress, G.** 2010. *Multimodality. A social semiotic approach to contemporary communication*. London: Routledge.
- Lawley, J.** 2015. "New software to help EFL students self-correct their writing". *Language Learning & Technology*, 19 (1), 23-33.
- Lawley, J.** 2016. "Spelling: computerised feedback for self-correction". *Computer Assisted Language Learning*, 29 (5), 868-880.
- Lee, I.** 2013. "Research into practice: Written corrective feedback". *Language Teaching*, 46 (1), 108-119.
- Long, M.** 1991. "Focus on form: A design feature in language teaching methodology". In de Bot, K., C. Kramsch and R. Ginsburg (Eds.) *Foreign language research in cross-cultural perspective*. Amsterdam: John Benjamins, 39-52.
- Lyster, R.** 1998. "Recasts, repetition and ambiguity in L2 classroom discourse". *Studies in Second Language Acquisition*, 20 (1), 51-80.
- Lyster, R. and Mori, H.** 2006. "Interactional feedback and instructional counterbalance". *Studies in Second Language Acquisition*, 28, 269-300.
- Mackey, A., Gass, S. and McDonough, K.** 2000. "How do learners perceive interactional feedback?". *Studies in Second Language Acquisition*, 22, 471-497.
- Master, P.** 1995. "Consciousness raising and article pedagogy". In Belcher, D. and G. Brain (Eds.) *Academic writing in a second language*. Norwood, NJ.: Ablex, 183-204.
- Nadasdi, T. and Sinclair, S.** 2007. *Anything I can do, CPU can do better: A comparison of human and computer grammar correction for L2 writing using BonPatron.com*. Unpublished manuscript. 15 January 2018. <https://sites.ualberta.ca/~tnadasdi/Dublin.htm>.

- Ortega, L.** 2009. "The linguistic environment". In Ortega, L. (Ed.) *Understanding second language acquisition*. London: Hodder Arnold, 71-76.
- Panova, I. and Lyster, R.** 2002. "Patterns of corrective feedback and uptake in an adult ESL classroom". *TESOL Quarterly*, 36, 573-595.
- Peterson, E. K.** 2017. "The impact of computer-generated feedback on student perceptions of revision process". *Masters of Arts in Education Action Research Papers*, 247. 28 August 2018. <https://sophia.stkate.edu/maed/247>
- Philp, J.** 2003. "Constraints on 'noticing the gap': Nonnative speakers' noticing of recasts in NS-NNS interaction". *Studies in Second Language Acquisition*, 25, 99-126.
- Polio, C., Fleck, C. and Leder, N.** 1998. "'If only I had more time': ESL learners' changes in linguistic accuracy on essay revisions". *Journal of Second Language Writing*, 7, 43-68.
- Potter, R. and Fuller, D.** 2008. "My new English partner? Using the Grammar Checker in writing instruction". *English Journal*, 98 (1), 36-41.
- Reiners, T., Dreher, C. and Dreher, H.** 2011. "Six key topics for automated assessment utilization and acceptance". *Informatics in Education*, 10 (1), 47-64.
- Russell, J. and Spada, N.** 2006. "The effectiveness of feedback for the acquisition of L2 grammar". In Norris, J. D. and L. Ortega (Eds.) *Synthesizing research on language learning and teaching*. Amsterdam: John Benjamins, 133-164.
- Sauro, S.** 2009. "Computer-mediated corrective feedback and the development of L2 grammar". *Language Learning & Technology*, 13 (1), 96-120.
- Schmidt, R.** 1990. "The role of consciousness in second language learning". *Applied Linguistics*, 11 (2), 129-158.
- Spada, N.** 2011. "Beyond form-focused instruction: Reflections on past, present and future research". *Language Teaching*, 44, 225-236.
- Tiene, D. and Luft, P.** 2001. "Teaching in a technology-rich classroom". *Educational Technology*, 41, 23-31.

- Truscott, J.** 1996. "The case against grammar correction in L2 writing classes". *Language Learning*, 46, 327-369.
- Van Beuningen, N., de Jong, N. H. and Kuiken, F.** 2008. "The effect of direct and indirect corrective feedback on L2 learners' written accuracy". *International Journal of Applied Linguistics*, 156, 279-296.
- Ware, P.** 2011. "Computer-generated feedback on student writing". *TESOL Quarterly*, 45 (4), 769-774.
- Ware, P. and Warschauer, M.** 2005. "Electronic feedback and second language writing". In Hyland, K. and F. Hyland (Eds.) *Feedback and second language writing*. Cambridge: Cambridge University Press, 1-29.
- Warschauer, M.** 2010. "Invited commentary: New tools for teaching writing". *Language Learning & Technology*, 14 (1), 3-8.

APPENDIX 1: Sample tailor-made test

1. She decided to write her grandma to ask for ~~an~~ advice.
2. Richard tried to make her feel comfortable and confident.
3. Grandma started another debate but with no success.
4. It was ^{the} a day of her 18th birthday when Pam wrote a letter to her grandma.
5. On ^{the} other hand ^{the} situation in Pam's home was very difficult.
6. ^{For} By this reason Pam wrote to grandma.
7. When she said her parents about this job they felt sadness.
8. In ^{the} letter Pam begged some money to her grandma to settle some bills.
9. In the end she managed to finish her studies and set up her own business.
10. Their parents were always motivating her to get ~~a~~ good marks.
11. Pam wrote to grandma to ask for some money ^{to} ~~to~~ her ambitious project.

Received: 15 March 2018

Accepted: 23 July 2018

Cite this article as:

Hernández Puertas, Tamara 2018. "Teacher's feedback vs. computer-generated feedback: A focus on articles". *Language Value* 10 (1), 68-89. Jaume I University ePress: Castelló, Spain. <http://www.e-revistas.uji.es/languagevalue>.

DOI: <http://dx.doi.org/10.6035/LanguageV.2018.10.5>

ISSN 1989-7103

Articles are copyrighted by their respective authors